A likelihood approximation for locally stationary processes

by Rainer Dahlhaus

Universität Heidelberg

Abstract

A new approximation to the Gaussian likelihood of a multivariate locally stationary process is introduced. It is based on an approximation of the inverse of the covariance matrix of such processes. The new quasi-likelihood is a generalisation of the classical Whittle-likelihood for stationary processes. For parametric models asymptotic normality and efficiency of the resulting estimator are proved. Since the likelihood has a special local structure it can be used for nonparametric inference as well. This is briefly sketched for different estimates.

1 Introduction

Suppose we observe data $X_1, \dots, X_T$ from some nonstationary process and we want to fit a parametric model to the data. An example is an autoregressive process with time varying coefficients where we model the coefficient functions by polynomials in time. If the process is Gaussian we can write down the exact likelihood function which, in the case of mean zero, takes the form

$$\begin{aligned} \mathcal{L}_T^{(e)}(\theta) &:= -\frac{1}{T} \text{ Gaussian log likelihood} \\ &= \frac{1}{2}\log(2\pi) + \frac{1}{2T}\log\det\Sigma_\theta + \frac{1}{2T}\underline{X}'\Sigma_\theta^{-1}\underline{X} \qquad (1.1) \end{aligned}$$

with $\underline{X} = (X_1, \dots, X_T)'$ (the assumption of a zero mean is given up later on).

However, for most of the time varying models the calculations needed for the minimisation of this function are too time consuming. Suppose for example we want to fit a time varying AR-model to the data where the coefficient functions are polynomials in

[0] *AMS 1991 subject classifications.* Primary 62M10; secondary 62F10.

[0] *Key words and phrases.* Locally stationary processes, Whittle-likelihood, local likelihood, preperiodogram, generalized Toeplitz matrices.

time and where the AR-model order and the polynomial orders have to be determined by a model selection criterion. Such a model can be written in state space form with time varying system matrices (cf. Dahlhaus, 1996b) and - in principle - the minimisation could be done as in the stationary case by using the prediction error decomposition, the Kalman filter and a numerical optimisation routine (cf. Harvey, 1989, Section 3.4). However, we usually have a high dimensional parameter space, time dependent system matrices and a large number of models at hand, which make the calculations practically impossible.

To overcome these problems we suggest in this paper an approximation to the above likelihood which is a generalisation of Whittle's approximation in the stationary case (cf. Whittle, 1953, 1954). In the stationary case Σ_θ is the Toeplitz matrix of the spectral density. Whittle had suggested to approximate Σ_θ^{-1} by the Toeplitz matrix of the inverse of the spectral density leading with the Szegö formula (cf. Grenander and Szegö, 1958, Section 5.2) to the Whittle likelihood

$$\mathcal{L}_T^{(W)}(\theta) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \log 4\pi^2 f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)} \right\} d\lambda$$

where

$$I_T(\lambda) = \frac{1}{2\pi T} \Big| \sum_{t=1}^{T} X_t \exp(-i\lambda t) \Big|^2$$

is the periodogram.

In this paper we derive a similar approximation for processes that only show locally some kind of stationary behaviour. More precisely we consider locally stationary processes as defined in Dahlhaus (1996a,b; 1997), i.e. processes with a time varying spectral representation as in (2.1) (the exact definition is given in Section 3). For an introduction to univariate locally stationary processes we refer to Dahlhaus (1996c).

In Section 2 we motivate the approximation and discuss its benefits in a simplified setting (univariate processes, mean zero). In Section 3 we introduce multivariate locally stationary processes and the generalisation of the Whittle likelihood for such processes. We then investigate the properties of the resulting parameter estimate.

Technically the approximation is based on a special generalisation of Toeplitz matrices (see (2.2)). The behaviour of norms and matrix products of such matrices is investigated in the appendix.

2 A motivation for the likelihood approximation

In this section we use a simplified setting to introduce and motivate the likelihood approximation and to discuss its applications. Furthermore, we compare it to the Whittle approximation in the stationary case.

Suppose the observed process has a time varying spectral representation of the form

$$X_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) A_\theta \left(\frac{t}{T}, \lambda\right) d\xi(\lambda) \qquad (t = 1, \ldots, T), \tag{2.1}$$

where $\xi(\lambda)$ is a stochastic process on $[-\pi, \pi]$ with mean zero and orthonormal increments. As e.g. in nonparametric regression the time parameter $u = t/T$ in A_θ is rescaled for a meaningful asymptotic theory (this is a special case of a locally stationary process as defined in Definition 3.1 below). We obtain for the variance covariance matrix Σ_θ

$$\begin{aligned} \Sigma_{\theta_{r,s}} &= \operatorname{cov}(X_{r,T}, X_{s,T}) \\ &= \int_{-\pi}^{\pi} \exp\{i\lambda(r-s)\} A_\theta \left(\frac{r}{T}, \lambda\right) \overline{A_\theta \left(\frac{s}{T}, \lambda\right)} d\lambda. \end{aligned}$$

In the stationary case where $A_\theta(\frac{r}{T}, \lambda) = A_\theta(\lambda)$ does not depend on time this is equal to

$$\int_{-\pi}^{\pi} \exp\{i\lambda(r-s)\} f_\theta(\lambda) d\lambda$$

where $f_\theta(\lambda) = |A_\theta(\lambda)|^2$ is the spectral density of the process. In the derivation of the Whittle approximation Σ_θ^{-1} is approximated by the Toeplitz matrix

$$\left\{ \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \exp\{i\lambda(r-s)\} f_\theta(\lambda)^{-1} d\lambda \right\}_{r,s=1,\ldots,T}.$$

In the nonstationary case we have for r, s close to each other and for a function A_θ which is smooth in time

$$\Sigma_{\theta_{r,s}} \approx \int_{-\pi}^{\pi} \exp(i\lambda(r-s)) f_\theta \left(\frac{r+s}{2T}, \lambda\right) d\lambda$$

where $f_\theta(u,\lambda) := |A_\theta(u,\lambda)|^2$ is the time-varying spectral density of the process.

This suggests to use now

$$\left\{ \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \exp\{i\lambda(r-s)\} f_\theta \left(\frac{r+s}{2T}, \lambda \right)^{-1} d\lambda \right\}_{r,s=1,\dots,T}$$

as an approximation of Σ_θ^{-1} in the nonstationary case. Since it leads to a slightly nicer criterion we use instead $U_T(\frac{1}{4\pi^2} f_\theta^{-1})$ where

$$U_T(\phi) = \left\{ \int_{-\pi}^{\pi} \exp\{i\lambda(r-s)\} \phi \left(\frac{1}{T} \left[\frac{r+s}{2} \right]^*, \lambda \right) d\lambda \right\}_{r,s=1,\dots,T} \tag{2.2}$$

and $[x]^*$ denotes the smallest integer larger or equal to x. Note that $U_T(\frac{1}{4\pi^2} f_\theta^{-1})$ is the classical Toeplitz/Whittle-approximation if f_θ is constant over time (stationary case). Using this approximation, i.e.

$$\Sigma_\theta^{-1} \approx U_T \left(\frac{1}{4\pi^2} f_\theta^{-1} \right)$$

and a generalization of Szegö's formula to the nonstationary case (see Proposition 3.4 below), nameley

$$\frac{1}{T} \log \det \Sigma_\theta \approx \frac{1}{2\pi} \int_0^1 \int_{-\pi}^{\pi} \log[2\pi f_\theta(u,\lambda)] d\lambda du$$

we obtain the following likelihood function as an approximation of the exact Gaussian likelihood $\mathcal{L}_T^{(e)}(\theta)$

$$\mathcal{L}_T^{(\ell)}(\theta) = \frac{1}{4\pi} \frac{1}{T} \sum_{t=1}^{T} \int_{-\pi}^{\pi} \log 4\pi^2 f_\theta \left(\frac{t}{T}, \lambda \right) d\lambda + \frac{1}{8\pi^2 T} \underline{X}' U_T(f_\theta^{-1}) \underline{X}.$$

The subsitution $[\frac{r+s}{2}]^* = t$, $r-s=k$ yields

$$\begin{aligned} \underline{X}' U_T(f_\theta^{-1}) \underline{X} &= \sum_{r,s=1}^{T} X_{r,T} X_{s,T} \int_{-\pi}^{\pi} \exp\{i\lambda(r-s)\} f_\theta \left(\frac{1}{T} \left[\frac{r+s}{2} \right]^*, \lambda \right)^{-1} d\lambda \\ &= \sum_{t=1}^{T} \int_{-\pi}^{\pi} f_\theta(\frac{t}{T}, \lambda)^{-1} \sum_k X_{[t+k/2],T} X_{[t-k/2],T} \exp(i\lambda k) d\lambda \end{aligned}$$

where the second sum is over all k such that $1 \le [t+k/2], [t-k/2] \le T$ and $[x]$ is the largest integer smaller or equal to x. Thus,

$$\mathcal{L}_T^{(\ell)}(\theta) = \frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}\Big\{\log 4\pi^2 f_\theta(\frac{t}{T},\lambda) + \frac{\tilde{I}_T(\frac{t}{T},\lambda)}{f_\theta(\frac{t}{T},\lambda)}\Big\}d\lambda \tag{2.3}$$

where

$$\tilde{I}_T(u,\lambda) := \frac{1}{2\pi}\sum_{k \atop 1\le[uT+k/2],[uT-k/2]\le T} X_{[uT+k/2],T}X_{[uT-k/2],T}\exp(i\lambda k).$$

$\tilde{I}_T(\frac{t}{T},\lambda)$ may be regarded as a local version of the periodogram at time t. It was introduced by Neumann and von Sachs (1997) as a starting point for a wavelet estimate of the time-varying spectral density. We will call $\tilde{I}_T(\frac{t}{T},\lambda)$ the preperiodogram at time t.

There exist several nice relations between the preperiodogram and the ordinary periodogram and the above likelihood and the Whittle-likelihood: We have

$$\begin{aligned} I_T(\lambda) &= \frac{1}{2\pi T}\Big|\sum_{r=1}^{T} X_{r,T}\exp(-i\lambda r)\Big|^2 \\ &= \frac{1}{2\pi}\sum_{k=-(T-1)}^{T-1}\frac{1}{T}\left(\sum_{t=1}^{T-|k|}X_{t,T}X_{t+|k|,T}\right)\exp(i\lambda k) \\ &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2\pi}\sum_{k \atop 1\le[t+k/2],[t-k/2]\le T} X_{[t+k/2],T}X_{[t-k/2],T}\exp(i\lambda k) \\ &= \frac{1}{T}\sum_{t=1}^{T}\tilde{I}_T(\frac{t}{T},\lambda), \end{aligned} \tag{2.4}$$

i.e. the periodogram is the average of the preperiodogram over time. (2.4) means that the periodogram $I_T(\lambda)$ is the Fourier transform of the covariance estimator of lag k over the whole segment while the preperiodogram $\tilde{I}_T(\frac{t}{T},\lambda)$ just uses the pair $X_{[t+k/2]}X_{[t-k/2]}$ as a kind of "local estimator" of the covariance of lag k at time t (note that $[t+k/2]-[t-k/2] = k$). For this reason Neumann and von Sachs also called $\tilde{I}_T(\frac{t}{T},\lambda)$ the localized periodogram.

A classical kernel estimator of the spectral density of a stationary process at some frequency λ_0 therefore can be regarded as an average of the preperiodogram over all time points and over the frequencies in the neighbourhood of λ_0. It is therefore plausible that averaging the preperiodogram around some frequency λ_0 and around some time-point t_0 gives an estimate of the time-varying spectrum $f(\frac{t_0}{T}, \lambda)$.

For a locally stationary process the preperiodogram is asymptotically unbiased. However, its variance explodes as T tends to infinity. Therefore, smoothing over time and frequency is essential to make a consistent estimate out of it. This smoothing is implicitely contained in the likelihood $\mathcal{L}_T^{(\ell)}(\theta)$. Instead of using $\tilde{I}_T(\frac{t}{T}, \lambda)$ in (2.3) one could think of using the classical periodogram over some small segment of data around t. Such a likelihood was studied in Dahlhaus (1997). The preperiodogram has advantages over such an estimate since a classical periodogram always contains some implicit smoothing over time (even if it is calculated over a small segment) which in the case of time varying spectra means that some information is getting lost. For this reason the preperiodogram is a valuable raw estimate, e.g. in (2.3) or for wavelet smoothing as in Neumann and von Sachs (1997). Another advantage in the context of Whittle estimation is that no segment length (e.g. for a periodogram) has to be selected.

The above likelihood $\mathcal{L}_T^{(\ell)}(\theta)$ coincides with the Whittle likelihood in the stationary case: If a stationary model is fitted, then $f_\theta(u, \lambda) = f_\theta(\lambda)$ is constant over time and the likelihood becomes

$$\begin{aligned}\mathcal{L}_T^{(\ell)}(\theta) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\{\log 4\pi^2 f_\theta(\lambda) + \frac{\frac{1}{T}\sum_{t=1}^{T}\tilde{I}_T(\frac{t}{T}, \lambda)}{f_\theta(\lambda)}\}d\lambda \\ &= \mathcal{L}_T^{(W)}(\theta).\end{aligned}$$

For that reason the results on the asymptotic behaviour of the minimizer of $\mathcal{L}_T^{(\ell)}(\theta)$ contain most of the results on the classical Whittle estimate as a special case (apart from our restriction to Gaussian processes). Among the large number of papers we mention the results of Dzhaparidze (1971) and Hannan (1973) for univariate time series, Dunsmuir (1979) for multivariate time series and Hosoya and Taniguchi (1982) for misspecified multivariate time series which follow as a special case from Theorem 3.8 below. A general overview over Whittle-estimates for stationary models may be found in the monograph of

Dzhaparidze (1986). We also mention the results of Klüppelberg and Mikosch (1996) on Whittle estimates for linear processes where the innovations have heavy tailed distributions, of Fox and Taqqu (1986) on Whittle estimates for long range dependent processes and of Robinson (1995) on semiparametric Whittle estimates for long range dependent processes. These results however are not a special case of Theorem 3.8.

There is another important aspect of the above likelihood approximation: The likelihood is of the form

$$\mathcal{L}_T^{(\ell)}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ell_T(\theta, \frac{t}{T})$$

with

$$\ell_T(\theta, \frac{t}{T}) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \{\log 4\pi^2 f_\theta(\frac{t}{T}, \lambda) + \frac{\tilde{I}_T(\frac{t}{T}, \lambda)}{f_\theta(\frac{t}{T}, \lambda)}\} d\lambda,$$

i.e. $\mathcal{L}_T^{(\ell)}(\theta)$ has a similar form as the negative log-likelihood function of iid observations where $\ell_T(\theta, \frac{t}{T})$ is the negative log-likelihood at time point t. In the present dependent situation $\ell_T(\theta, \frac{t}{T})$ may still be regarded as the negative log-likelihood at time point t which now in addition contains the full information on the dependence (correlation) structure of $X_{t,T}$ with all the other variables.

To illustrate this we give two examples:

1. Suppose we have the situation of nonparametric regression with heteroscedastic errors, i.e. our model is

$$X_{t,T} = m(\frac{t}{T}) + \sigma(\frac{t}{T})\varepsilon_t, \qquad \varepsilon_t \text{ iid } \mathcal{N}(0,1),$$

with $m(u) = m_\theta(u)$, $\sigma(u) = \sigma_\theta(u)$. This process is locally stationary in the sense of Definition 3.1 below. Since the mean is different from zero, the preperiodogram in $\ell(\theta, \frac{t}{T})$ contains an extra term (see (3.7) below). It is easy to show that in this case

$$\ell_T(\theta, \frac{t}{T}) = \frac{1}{2} \log 2\pi\sigma_\theta^2(\frac{t}{T}) + \frac{1}{2\sigma_\theta^2(\frac{t}{T})}(X_{t,T} - m_\theta(\frac{t}{T}))^2$$

which is exactly the Gaussian log-likelihood.

2. Suppose

$$X_{t,T} = a(\frac{t}{T})X_{t-1,T} + \sigma(\frac{t}{T})\varepsilon_t, \qquad \varepsilon_t \text{ iid } \mathcal{N}(0,1),$$

with $a(u) = a_\theta(u)$, $\sigma(u) = \sigma_\theta(u)$. Then $X_{t,T}$ is locally stationary with time varying spectrum

$$f_\theta(u,\lambda) = \frac{\sigma_\theta^2(u)}{2\pi}|1 - a_\theta(u)e^{i\lambda}|^{-2}$$

leading to

$$\ell_T(\theta, \frac{t}{T}) = \frac{1}{2}\log 2\pi\sigma_\theta^2(\frac{t}{T}) + \frac{1}{2\sigma_\theta^2(\frac{t}{T})}(X_{t,T} - a_\theta(\frac{t}{T})X_{t-1,T})^2 + r_t$$

with

$$r_t = a(\frac{t}{T})^2(X_{t,T}^2 - X_{t-1,T}^2),$$

i.e. $\sum_{t=1}^T r_t = O_p(1)$.

The fact that $\ell_T(\theta, \frac{t}{T})$ can be seen as the local likelihood of the process at time t opens the door for various nonparametric estimation methods. In this situation the model is parametrized by one or several curves in time (eg. as in Example 3.2).

Recall that several nonparametric estimation techniques can be written as the solution of a least squares problem, for example for the simple nonparametric regression problem

$$X_{t,T} = m(\frac{t}{T}) + \varepsilon_t$$

a) a kernel estimate can be written as

$$\hat{m}(u) = \underset{m}{\operatorname{argmin}} \frac{1}{b_T T}\sum_t K\left(\frac{u - t/T}{b_T}\right)\{X_{t,T} - m\}^2$$

where K is the kernel and b_T is some bandwidth;

b) a local polynomial fit can be written as

$$\hat{c}(u) = \underset{c}{\operatorname{argmin}} \frac{1}{b_T T} \sum_t K\left(\frac{u - t/T}{b_T}\right) \left\{ X_{t,T} - \sum_{j=0}^{d} c_j (\frac{t}{T} - u)^j \right\}^2;$$

where $c = (c_0, \ldots, c_d)'$ are the coefficients of the fitted polynomial at time u;

c) an orthogonal series estimator (e.g. wavelets) can be written as

$$\bar{\alpha} = \underset{\alpha}{\operatorname{argmin}} \frac{1}{T} \sum_t \left\{ X_{t,T} - \sum_{j=1}^{J} \alpha_j \psi_j (\frac{t}{T}) \right\}^2$$

together with some shrinkage to obtain the final estimator $\hat{\alpha}$. Here the $\psi_j(\cdot)$ $(j = 1, \ldots, J)$ denote some orthonormal functions. J usually increases with T.

Note that the $\{\ldots\}$-brackets always contain the negative log likelihood of the parameters up to some constants.

Suppose now we have a locally stationary model which is parametrized by one or several curves in time. By using the local likelihood we may define completely analogous to above

a) a kernel estimate by

$$\hat{\theta}(u) = \underset{\theta}{\operatorname{argmin}} \frac{1}{b_T T} \sum_{t=1}^{T} K\left(\frac{u - t/T}{b_T}\right) \ell_T\left(\theta, \frac{t}{T}\right);$$

b) a local polynomial fit by

$$\hat{c}(u) = \underset{c}{\operatorname{argmin}} \frac{1}{b_T T} \sum_{t=1}^{T} K\left(\frac{u - t/T}{b_T}\right) \ell_T\left(\sum_{j=0}^{d} c_j (\frac{t}{T} - u)^j, \frac{t}{T}\right);$$

c) an orthogonal series estimator (e.g. wavelets) by

$$\bar{\alpha} = \underset{\alpha}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} \ell_T\left(\sum_{j=1}^{J} \alpha_j \psi_j (\frac{t}{T}), \frac{t}{T}\right);$$

together with some shrinkage of $\bar{\alpha}$.

In case of several parameter curves (a vector of curves) θ, the c_j and the α_j are also vectors. In case of a multivariate process or a process with mean different from zero the definition (3.6) of $\ell_T(\theta, \frac{t}{T})$ has to be used.

It is obvious that the properties of these estimators have to be investigated in detail. However, this is quite complicated and would exceed the scope of this paper. We only want to demonstrate that the likelihood representation may have important applications in nonparametric estimation as well.

In the next section we prove that $\mathcal{L}_T^{(\ell)}(\theta)$ indeed is a good approximation of the exact Gaussian likelihood $\mathcal{L}_T^{(e)}(\theta)$. Furthermore, we consider parametric models and prove that the resulting parameter estimates are consistent, asymptotically normal and efficient. We do this for a larger class of processes than discussed in this section. In particular, we study multivariate locally stationary processes and allow the mean to be a function different from zero which introduces extra terms into the above expressions.

3 Asymptotic properties of parameter estimates

We start with the definition of a multivariate locally stationary process.

(3.1) Definition A sequence of multivariate stochastic processes $X_{t,T} = (X_{t,T}^{(1)}, ..., X_{t,T}^{(d)})'$ $(t = 1, ..., T)$ is called locally stationary with transfer function matrix A^o and mean function vector μ if there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,T}^o(\lambda) d\xi(\lambda) \tag{3.1}$$

where

(i) $\xi(\lambda)$ is a stochastic vector process on $[-\pi, \pi]$ with $\overline{\xi_a(\lambda)} = \xi_a(-\lambda)$ and

$$\operatorname{cum}\{d\xi_{a_1}(\lambda_1), ..., d\xi_{a_k}(\lambda_k)\} = \eta\left(\sum_{j=1}^{k} \lambda_j\right) h_{a_1 ... a_k}(\lambda_1, ..., \lambda_{k-1}) d\lambda_1 ... d\lambda_k$$

where cum$\{...\}$ denotes the cumulant of $k-th$ order, $h_a = 0$, $h_{ab}(\lambda) = \delta_{ab}$, $|h_{a_1 \ldots a_k}(\lambda_1, ..., \lambda_{k-1})| \leq \text{const}_k$ for all $a_1, ..., a_k \in \{1, ..., d\}$ and $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period 2π extension of the Dirac delta function.

(ii) There exists a constant K and a 2π-periodic matrix valued function $A : [0,1] \times \mathbb{R} \to \mathbb{C}^{d \times d}$ with $\overline{A(u,\lambda)} = A(u, -\lambda)$ and

$$\sup_{t,\lambda} |A^o_{t,T}(\lambda)_{ab} - A\left(\frac{t}{T}, \lambda\right)_{ab}| \leq KT^{-1} \tag{3.2}$$

for all $a, b = 1, \ldots, d$ and $T \in \mathbb{N}$. $A(u, \lambda)$ and $\mu(u)$ are assumed to be continuous in u.

$f(u, \lambda) := A(u, \lambda)\overline{A(u, \lambda)}'$ is the time varying spectral density matrix of the process.

Processes with an evolutionary spectral representation were introduced and investigated by Priestley (1965, 1981). The above definition is the multivariate generalization of the definition of univariate local stationarity as given in Dahlhaus (1997). This approach to local stationarity may be regarded as a setting which allows for a meaningful asymptotic theory for processes with an evolutionary spectral representation. The classical asymptotics for stationary sequences is contained as a special case (if μ and A do not depend on t). A detailed discussion of this definition and a comparison to Priestley's approach can be found in Dahlhaus (1996c). Another definition of local stationarity has recently been given by Mallat, Papanicolaou and Zhang (1998). We remark that the methods presented in this paper do not depend on the special definition of local stationarity. In some sense the above definition is only a framework for investigating the asymptotic properties of the estimates.

Examples of locally stationary processes in the univariate case can be found in Dahlhaus (1996a). For the multivariate case we give the following examples.

(3.2) Examples (i) Suppose Y_t is a multivariate stationary process, $\mu(\cdot)$ is a vector function and $\Sigma(\cdot)$ is a matrix function. Then

$$X_{t,T} = \mu(\frac{t}{T}) + \Sigma(\frac{t}{T})Y_t$$

is locally stationary. If Y_t is an iid sequence we have the situation of multivariate non-parametric regression.

(ii) Suppose $X_{t,T}$ is a time varying multivariate ARMA-model, that is $X_{t,T}$ is definied by the difference equations

$$\sum_{j=0}^{p} \Phi_j(\frac{t}{T}) \left(X_{t-j,T} - \mu(\frac{t-j}{T}) \right) = \sum_{j=0}^{q} \Psi_j(\frac{t}{T}) \Sigma(\frac{t-j}{T}) \varepsilon_{t-j}$$

where ε_t are iid with mean zero and variance-covariance matrix I_d and $\Phi_o(u) \equiv \Psi_o(u) \equiv I_d$. Under regularity conditions on the coefficient functions $\Phi_j(u)$ and $\Psi_j(u)$ it can be shown similarly to the univariate case (Dahlhaus, 1996a, Theorem 2.3) that these difference equations define a locally stationary process of the form (3.1). The time varying spectral density of the process is

$$f(u,\lambda) = \frac{1}{2\pi} \Phi(u,\lambda)^{-1} \Psi(u,\lambda) \Sigma(u) \Psi(u,-\lambda)' \Phi(u,-\lambda)'^{-1}$$

where $\Phi(u,\lambda) = \sum_{j=0}^{p} \Phi_j(u) e^{i\lambda j}$ and $\Psi(u,\lambda) = \sum_{j=0}^{q} \Psi_j(u) e^{i\lambda j}$. We omit details of the derivation. However, we remark that in this case the functions $A^o_{t,T}(\lambda)$ and $A(t/T,\lambda)$ do not coincide. They only fulfill (3.2).

In the following we look at parametric locally stationary models. An example is the case where the curves in the above examples are parametrized in time, e.g. by polynomials (for an example see Dahlhaus, 1997, Section 6).

Let $\underline{X} = (X'_{1,T}, \ldots, X'_{T,T})'$, $\underline{\mu} = (\mu(\frac{1}{T})', \ldots, \mu(\frac{T}{T})')'$, and let the $dT \times dT$-matrices $\Sigma_T(A,B)$ and $U_T(\phi)$ be defined by

$$\Sigma_T(A,B)_{r,s} = \int_{-\pi}^{\pi} \exp(i\lambda(r-s)) A^o_{r,T}(\lambda) B^o_{s,T}(-\lambda)' d\lambda \tag{3.3}$$

and

$$U_T(\phi)_{r,s} = \int_{-\pi}^{\pi} \exp(i\lambda(r-s)) \phi \left(\frac{1}{T} \left[\frac{r+s}{2} \right]^*, \lambda \right) d\lambda \tag{3.4}$$

$(r,s=1,\dots T)$ where $A^o_{r,T}(\lambda)$, $B^o_{r,T}(\lambda)$ and $\phi(u,\lambda)$ are $d\times d$-matrices. Then the exact Gaussian likelihood is

$$\mathcal{L}_T^{(e)}(\theta) := \frac{d}{2}\log(2\pi) + \frac{1}{2T}\log\det\Sigma_\theta + \frac{1}{2T}(\underline{X}-\underline{\mu}_\theta)'\Sigma_\theta^{-1}(\underline{X}-\underline{\mu}_\theta) \tag{3.5}$$

where $\Sigma_\theta = \Sigma_T(A_\theta, A_\theta)$ and $E(\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})' = \Sigma_T(A,A)$ with A from Definition 3.1.

We now proceed as in the univariate case (Section 2) to find a local likelihood approximation. We use a generalisation of the multivariate Szegö identity (see Proposition 3.4 below) and $U_T(\frac{1}{4\pi^2}f_\theta^{-1})$ as an approximation of Σ_θ^{-1} to obtain

$$\begin{aligned}\mathcal{L}_T(\theta) \;&:=\; \mathcal{L}_T^{(\ell)}(\theta) := \frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}\log[(2\pi)^{2d}\det f_\theta(\frac{t}{T},\lambda)]d\lambda \\ &\qquad\qquad + \frac{1}{8\pi^2T}(\underline{X}-\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta) \\ &=\; \frac{1}{T}\sum_{t=1}^{T}\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log\left[(2\pi)^{2d}\det f_\theta(\frac{t}{T},\lambda)\right] + \operatorname{tr}\left[f_\theta(\frac{t}{T},\lambda)^{-1}\tilde{I}_T^{\mu_\theta}(\frac{t}{T},\lambda)\right]\right\}d\lambda \\ &=:\; \frac{1}{T}\sum_{t=1}^{T}\ell_T(\theta,\frac{t}{T}) \end{aligned}\tag{3.6}$$

where

$$\begin{aligned}\tilde{I}_T^{\mu}(u,\lambda)_{ab} := \frac{1}{2\pi}\sum_{k\,:\,1\le[uT+k/2],[uT-k/2]\le T}&\left[X^{(a)}_{[uT+k/2],T} - \mu^{(a)}\left(\frac{[uT+k/2]}{T}\right)\right] \\ &\times\left[X^{(b)}_{[uT-k/2],T} - \mu^{(b)}\left(\frac{[uT-k/2]}{T}\right)\right]\exp(-i\lambda k)\end{aligned}\tag{3.7}$$

is the multivariate version of the preperiodogram.

In the univariate case and for $\mu_\theta = 0$ this is the likelihood we have already discussed in Section 2. We call $\ell_T(\theta,\frac{t}{T})$ the local likelihood at time t. If the mean is not zero and one is not interested in modelling the mean one may use $\tilde{I}_T^{\hat{\mu}}(u,\lambda)$ instead of $\tilde{I}_T^{\mu_\theta}(u,\lambda)$ where $\hat{\mu}$ is the arithmetic mean or some kernel estimate (if the mean is not believed to be constant over time).

Before investigating the asymptotic properties of the minimizer of $\mathcal{L}_T(\theta)$ we prove some results on the likelihood approximation itself. First we state two results which show that $U_T(\{4\pi^2 f_\theta\}^{-1})$ and $\mathcal{L}_T(\theta)$ are approximations of Σ_θ^{-1} and $\mathcal{L}_T^{(e)}$ respectively. We also show that

$$\begin{aligned}\mathcal{L}(\theta) \quad &:= \quad \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{\log\det f_\theta(u,\lambda)+\operatorname{tr}[f_\theta(u,\lambda)^{-1}f(u,\lambda)]\right\}d\lambda du \\ &\qquad + \frac{1}{4\pi}\int_0^1(\mu_\theta(u)-\mu(u))'f_\theta^{-1}(u,0)(\mu_\theta(u)-\mu(u))du \qquad (3.8)\end{aligned}$$

is the limit of $\mathcal{L}_T(\theta)$ and $\mathcal{L}_T^{(e)}(\theta)$.

The technical parts of the following proofs consist of the derivation of properties of products of matrices $\Sigma_T(A,B)$, $\Sigma_T(A,A)^{-1}$ and $U_T(\phi)$. These properties are derived in the appendix. In particular Lemma A.1, A.5 and A.8 are of relevance for the following proofs.

For convenience we refer in the following proposition to Assumption A.3 in the appendix concerning the smoothness of the transfer function and the mean. These conditions are fulfilled under Assumption 3.6 below. By $\|A\|$ and $\left|\!\left|\!\left| A \right|\!\right|\!\right|$ we denote the spectral norm and the Euclidean norm of a matrix A (cp. (A.1) and (A.2)). $\|v\|_2$ is the Euclidean norm of a vector.

(3.3) Proposition *Suppose the matrices A and ϕ fulfill the smoothness conditions of Assumption A.3 (i) - (iii) (appendix) with existing and bounded derivatives $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial\lambda}A(u,\lambda)_{ab}$ and eigenvalues of $\phi(u,\lambda)$ which are bounded from below uniformly in u and λ. Then we have*

$$\frac{1}{T}\left|\!\left|\!\left| \Sigma_T(A,A)^{-1} - U_T(\{4\pi^2 A\bar{A}'\}^{-1}) \right|\!\right|\!\right|^2 = O(T^{-1}\ln^3 T) \qquad (3.9)$$

and

$$\frac{1}{T}\left|\!\left|\!\left| U_T(\phi)^{-1} - U_T(\{4\pi^2\phi\}^{-1}) \right|\!\right|\!\right|^2 = O(T^{-1} ln^{-23} T).$$

Proof. Let $\Sigma_T = \Sigma_T(A,A)$ and $U_T = U_T(\{4\pi^2 A\bar{A}'\}^{-1})$. We obtain with Lemma A.1 (b,c) and Lemma A.5

$$\begin{aligned}\frac{1}{T}\left|\!\left|\!\left| \Sigma_T^{-1} - U_T \right|\!\right|\!\right|^2 \quad &\leq \quad \frac{1}{T}\left|\!\left|\!\left| I - \Sigma_T^{1/2}U_T\Sigma_T^{1/2} \right|\!\right|\!\right|^2 \|\Sigma_T^{-1}\|^2 \\ &\leq \quad K\Big(d - \frac{2}{T}\operatorname{tr}\{U_T\Sigma_T\} + \frac{1}{T}\operatorname{tr}\{U_T\Sigma_T U_T\Sigma_T\}\Big).\end{aligned}$$

Lemma A.7 (i) now implies the result. The second result is obtained in the same way with Lemma A.8. $\square$

We now state the generalisation of the Szegö identity (cf. Grenander and Szegö, 1958, Section 5.2) to multivariate locally stationary process.

(3.4) Proposition *Suppose A fulfills Assumption A.3 (i), (ii), with bounded derivatives $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial \lambda}A(u,\lambda)_{ab}$. Then we have with $f(u,\lambda) = A(u,\lambda)A(u,-\lambda)'$*

$$\frac{1}{T}\log\det\Sigma_T(A,A) = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi}\log[(2\pi)^d\det f(u,\lambda)]d\lambda du + O(T^{-1}\ln^{11}T).$$

If $A = A_\theta$ depends on a parameter θ and fulfills the smoothness conditions of Assumption 3.6 (iii), (iv), then the $O(T^{-1}\ln^{11}T)$ term is uniform in θ.

Proof. The proof can be found in A.9 of the appendix. $\square$

From now on we set $\nabla_i = \frac{\partial}{\partial\theta_i}$ and $\nabla_{ij}^2 = \frac{\partial^2}{\partial\theta_i\partial\theta_j}$.

(3.5) Theorem *Suppose $X_{t,T}$ is a locally stationary Gaussian process with transfer function matrix A^o and mean function vector μ and we fit a locally stationary model with transfer function matrix A_θ^o and mean function vector μ_θ. Suppose further that all eigenvalues of $f_\theta(u,\lambda) = A_\theta(u,\lambda)\overline{A_\theta(u,\lambda)}'$ are bounded from below uniformly in u and λ and the components of A, A_θ, μ, μ_θ are differentiable with uniformly bounded derivatives $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial\lambda}A(u,\lambda)_{ab}$, $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial\lambda}A_\theta(u,\lambda)_{ab}$, $\frac{\partial}{\partial u}\mu(u)_a$, $\frac{\partial}{\partial u}\mu_\theta(u)_a$ respectively. Then we have*

(i)

$$\mathcal{L}_T(\theta) - \mathcal{L}_T^{(e)}(\theta) = O_P(T^{-1}\ln^{11}T).$$

(ii) If in addition the first derivatives $\nabla_j A_\theta(u,\lambda)_{ab}$ and $\nabla_j\mu_\theta(u)_a$ fulfill the above smoothness properties, we also have

$$\nabla_j\mathcal{L}_T(\theta) - \nabla_j\mathcal{L}_T^{(e)}(\theta) = O_P(T^{-1}\ln^{23}T).$$

(iii) Furthermore,

$$\mathcal{L}(\theta) = \lim_{T\to\infty} E\mathcal{L}_T^{(e)}(\theta) = \lim_{T\to\infty} E\mathcal{L}_T(\theta)$$

and

$$\mathcal{L}_T^{(e)}(\theta) \xrightarrow{P} \mathcal{L}(\theta), \quad \mathcal{L}_T(\theta) \xrightarrow{P} \mathcal{L}(\theta).$$

A similar result also holds for the higher order derivatives of the likelihoods. We conjecture that also a uniform result (in θ) holds and that the log-terms and the Gaussian assumption can be dropped. However, a uniform result requires much more effort. In order not to blow up the paper we omit these generalisations.

PROOF. (i) We obtain with Proposition 3.4 and $B_T := \Sigma_T(A_\theta, A_\theta)^{-1} - U_T(\{4\pi^2 A_\theta \bar{A}_\theta'\}^{-1})$

$$\mathcal{L}_T(\theta) - \mathcal{L}_T^{(e)}(\theta) = \frac{1}{2T}(\underline{X} - \underline{\mu}_\theta)' B_T (\underline{X} - \underline{\mu}_\theta) + O(T^{-1}\ln^{11} T).$$

Since

$$\begin{aligned} & \frac{1}{T}(\underline{X} - \underline{\mu}_\theta) B_T (\underline{X} - \underline{\mu}_\theta) \\ = \ & \frac{1}{T}(\underline{X} - \underline{\mu}) B_T (\underline{X} - \underline{\mu}) \\ + \ & \frac{2}{T}(\underline{X} - \underline{\mu})' B_T (\underline{\mu} - \underline{\mu}_\theta) \\ + \ & \frac{1}{T}(\underline{\mu} - \underline{\mu}_\theta)' B_T (\underline{\mu} - \underline{\mu}_\theta) \end{aligned} \tag{3.10}$$

we obtain with Lemma A.8 and $\Sigma = \Sigma_T(A, A)$

$$\begin{aligned} E\{\mathcal{L}_T(\theta) - \mathcal{L}_T^{(e)}(\theta)\} &= \frac{1}{2T}\mathrm{tr}\{B_T\Sigma\} + \frac{1}{2T}(\underline{\mu} - \underline{\mu}_\theta)' B_T (\underline{\mu} - \underline{\mu}_\theta) + O(T^{-1}\ln^{11} T) \\ &= O(T^{-1}\ln^{11} T) \end{aligned}$$

and

$$\begin{aligned} \mathrm{var}\{\mathcal{L}_T(\theta) - \mathcal{L}_T^{(e)}(\theta)\} &= \frac{1}{2T^2}\mathrm{tr}\{B_T\Sigma B_T\Sigma\} + \frac{1}{T^2}(\underline{\mu} - \underline{\mu}_\theta)' B_T \Sigma B_T (\underline{\mu} - \underline{\mu}_\theta) \\ &= O(T^{-2}\ln^{23} T) \end{aligned}$$

which implies the result.

(ii) We obtain with Lemma A.8, B_T as above and $C_T := -\Sigma_T(A_\theta, A_\theta)^{-1}\{\Sigma_T(\nabla_j A_\theta, A_\theta) + \Sigma_T(A_\theta, \nabla_j A_\theta)\}\Sigma_T(A_\theta, A_\theta)^{-1} - U_T(\nabla_j\{4\pi^2 A_\theta \bar{A}'_\theta\}^{-1})$

$$\nabla_j \mathcal{L}_T(\theta) - \nabla_j \mathcal{L}_T^{(e)}(\theta) = \frac{1}{2T}(\underline{X} - \underline{\mu}_\theta)' C_T(\underline{X} - \underline{\mu}_\theta) - \frac{1}{T}(\nabla_j \underline{\mu}_\theta)' B_T(\underline{X} - \underline{\mu}_\theta) + O(T^{-1}\ln^{11} T).$$

Analogously to above we obtain with Lemma A.8

$$E(\nabla_j \mathcal{L}_T(\theta) - \nabla_j \mathcal{L}_T^{(e)}(\theta)) = O(T^{-1}\ln^{23} T)$$

and

$$\mathrm{var}(\nabla_j \mathcal{L}_T(\theta) - \nabla_j \mathcal{L}_T^{(e)}(\theta)) = O(T^{-2}\ln^{47} T)$$

which gives the result.

(iii) follows similarly to (i) (use e.g. $B_T = \Sigma_T(A_\theta, A_\theta)^{-1}$ in the above derivation). $\square$

Theorem 3.5 (iii) basically gives the asymptotic Kullback-Leibler-information divergence of two multivariate locally stationary processes: If $X_{t,T}(\tilde{X}_{t,T})$ are multivariate locally stationary with spectral densities $f = A\bar{A}'(\tilde{f} = \tilde{A}\bar{\tilde{A}}')$, mean functions $\mu(\tilde{\mu})$ and Gaussian densities $g(\tilde{g})$, then we obtain for the information divergence

$$\begin{aligned} D(\tilde{f}, \tilde{\mu}, f, \mu) &= \lim_{T\to\infty} \frac{1}{T} E_g \log \frac{g}{\tilde{g}} \\ &= \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \{\log\det[\tilde{f}(u,\lambda) f(u,\lambda)^{-1}] + \mathrm{tr}[\tilde{f}(u,\lambda)^{-1} f(u,\lambda) - I]\} d\lambda du \\ &\quad + \frac{1}{4\pi}\int_0^1 (\tilde{\mu}(u) - \mu(u))' \tilde{f}(u,0)^{-1}(\tilde{\mu}(u) - \mu(u)) du. \end{aligned}$$

This is the time average of the Kullback-Leibler divergence in the stationary case (cf. Parzen, 1983, for the univariate stationary case with mean zero).

We now study the behaviour of

$$\hat{\theta}_T := \underset{\theta\in\Theta}{\mathrm{argmin}}\, \mathcal{L}_T(\theta).$$

Furthermore, let

$$\theta_0 := \operatorname*{argmin}_{\theta \in \Theta} \mathcal{L}(\theta).$$

The results are proved under the following assumptions.

(3.6) Assumption (i) We observe a realisation $X_{1,T}, ..., X_{T,T}$ of a d-dimensional stationary Gaussian process with true mean function vector μ and transfer function matrix A^o and fit a class of locally stationary Gaussian processes with mean function vector μ_θ and transfer function matrix A^o_θ, $\theta \in \Theta \subset \mathbb{R}^p$, Θ compact.

(ii) $\theta_0 = \operatorname{argmin} \mathcal{L}(\theta)$ exists uniquely and lies in the interior of Θ.

(iii) The components of $A_\theta(u, \lambda)$ are differentiable in θ, u and λ with uniformly continuous derivatives $\nabla^2_{ij} \frac{\partial^2}{\partial u^2} \frac{\partial}{\partial \lambda} A_\theta(u, \lambda)_{ab}$.

(iv) All eigenvalues of $f_\theta(u, \lambda) = A_\theta(u, \lambda) A_\theta(u, -\lambda)'$ are bounded from below by some constant $C > 0$ uniformly in θ, u and λ.

(v) The components of $A(u, \lambda)$ are differentiable in u and λ with uniformly bounded derivatives $\frac{\partial}{\partial u} \frac{\partial}{\partial \lambda} A(u, \lambda)_{ab}$.

(vi) The components of $\mu(u)$, $\mu_\theta(u)$, $\nabla_i \mu_\theta(u)$ and $\nabla^2_{ij} \mu_\theta(u)$ are differentiable in u with uniformly bounded derivatives.

In the case where the model is correctly specified, i.e. $A(u, \lambda) = A_{\theta^*}(u, \lambda)$ and $\mu(u) = \mu_{\theta^*}(u)$ with some $\theta^* \in \Theta$ one can show that $\theta_0 = \theta^*$.

(3.7) Theorem *Suppose that Assumption 3.6 holds. Then*

$$\hat{\theta}_T \xrightarrow{P} \theta_0.$$

Proof. The basic idea is taken from Walker (1964), Section 2. In Theorem 3.5 (iii) we have proved that

$$\mathcal{L}_T(\theta) \xrightarrow{P} \mathcal{L}(\theta).$$

Since θ_0 is assumed to be unique it follows that for all $\theta_1 \neq \theta_0$ there exists a constant $c(\theta_1) > 0$ with

$$\lim_{T\to\infty} P(\mathcal{L}_T(\theta_1) - \mathcal{L}_T(\theta_0) < c(\theta_1)) = 0.$$

Furthermore, we have with a mean value $\bar{\theta}$

$$\mathcal{L}_T(\theta_2) - \mathcal{L}_T(\theta_1) = (\theta_2 - \theta_1)'\nabla\mathcal{L}_T(\bar{\theta})$$

where (cp. (3.10))

$$\begin{aligned}
\nabla_i\mathcal{L}_T(\theta) &= \frac{1}{4\pi}\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}\mathrm{tr}\left\{f_\theta(\frac{t}{T},\lambda)\nabla_i f_\theta(\frac{t}{T},\lambda)^{-1}\right\}d\lambda \\
&\quad +\frac{1}{8\pi^2 T}(\underline{X}-\underline{\mu}_\theta)'U_T(\nabla_i f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta) \qquad (3.11)\\
&\quad -\frac{1}{4\pi^2 T}(\nabla_i\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta) \\
&= \frac{1}{8\pi^2 T}(\underline{X}-\underline{\mu})'U_T(\nabla_i f_\theta^{-1})(\underline{X}-\underline{\mu}) \\
&\quad +\frac{1}{4\pi^2 T}\nabla_i\left\{(\underline{\mu}-\underline{\mu}_\theta)'U_T(f_\theta^{-1})\right\}(\underline{X}-\underline{\mu}) + \mathrm{const.} \qquad (3.12)
\end{aligned}$$

with a constant independent of $\underline{X}$ (but dependent on θ and T). With the Cauchy-Schwarz inequality and Lemma A.1(h) we get

$$\begin{aligned}
&\frac{1}{T}(\nabla_i\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta) \\
\leq\ &\frac{1}{T}\left\{(\nabla_i\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\nabla_i\underline{\mu}_\theta)\cdot(\underline{X}-\underline{\mu}_\theta)'U_T(f_\theta^{-1})(\underline{X}-\underline{\mu}_\theta)\right\}^{1/2} \\
\leq\ &\left\{\frac{1}{T}\|\nabla_i\mu_\theta\|_2^2\right\}^{1/2}\left\{\frac{2}{T}\|\underline{X}\|_2^2+\frac{2}{T}\|\underline{\mu}_\theta\|_2^2\right\}^{1/2}\|U_T(f_\theta^{-1})\|
\end{aligned}$$

which by Assumption 3.6 and Lemma A.5 is uniformly bounded by

$$K + K\frac{1}{T}\|\underline{X}\|_2^2.$$

Similarly, we can estimate the other terms in (3.11) leading to

$$\sup_{\theta_2 \in U_\delta(\theta_1)} |\mathcal{L}_T(\theta_2) - \mathcal{L}_T(\theta_1)| \leq K\delta(1 + \frac{1}{T}\underline{X}'\underline{X})$$

with some constant K. Since $E\frac{1}{T}\underline{X}'\underline{X} = \frac{1}{T}\|\underline{\mu}\|_2^2 \to \sum_{a=1}^T \int_0^1 \mu_a(u)^2 du$ and $\text{var}\frac{1}{T}\underline{X}'\underline{X} = \frac{2}{T^2}\text{tr}\{\Sigma^2\} \leq \frac{2}{T}\|\Sigma\|^2 \leq \frac{K}{T}$ (Lemma A.1 and A.5) $T^{-1}\underline{X}'X$ is bounded in probability. Thus there exists for all $\theta_1 \neq \theta_0$ a $c(\theta_1) > 0$ and $a\ \delta = \delta(\theta_1)$ with

$$\begin{aligned}
&\lim_{T\to\infty} P(\inf_{\theta_2 \in U_\delta(\theta_1)} \mathcal{L}_T(\theta_2) - \mathcal{L}_T(\theta_0) \geq c(\theta_1)/2) \\
&\quad \geq \ 1 - \lim_{T\to\infty} P(\mathcal{L}_T(\theta_1) - \mathcal{L}_T(\theta_0) < c(\theta_1)) - \lim_{T\to\infty} P(\sup_{\theta_2 \in U_\delta(\theta_1)} |\mathcal{L}_T(\theta_2) - \mathcal{L}_T(\theta_1)| \geq c(\theta_1)/2) \\
&\quad = \ 1.
\end{aligned}$$

A compactness argument as in Walker (1964) implies the result. $\square$

(3.8) Theorem *Suppose that Assumption 3.6 holds. Then we have*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma^{-1}V\Gamma^{-1})$$

with

$$\begin{aligned}
\Gamma_{ij} \ &= \ \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \text{tr}\left\{(f - f_{\theta_0})\nabla_{ij} f_{\theta_0}^{-1}\right\} d\lambda du - \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \text{tr}\left\{(\nabla_i f_{\theta_0})(\nabla_j f_{\theta_0}^{-1})\right\} d\lambda du \\
&\quad + \frac{1}{4\pi}\int_0^1 \nabla_{ij}^2 \left\{(\mu(u) - \mu_{\theta_0}(u))' f_{\theta_0}^{-1}(u,0)(\mu(u) - \mu_{\theta_0}(u))\right\} du
\end{aligned}$$

and

$$\begin{aligned}
V_{ij} \ &= \ \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \text{tr}\left\{f\,(\nabla_i f_\theta^{-1}) f\,(\nabla_j f_\theta^{-1})\right\} d\lambda du \\
&\quad + \frac{1}{2\pi}\int_0^1 \int_{-\pi}^{\pi} \left[\nabla_i \left\{(\mu(u) - \mu_{\theta_0}(u))' f_{\theta_0}^{-1}(u,0)\right\}\right] f(u,0) \left[\nabla_j \left\{f_{\theta_0}^{-1}(u,0)(\mu(u) - \mu_{\theta_0}(u))\right\}\right] du.
\end{aligned}$$

Proof. We obtain with the mean value theorem

$$\nabla_i \mathcal{L}_T(\hat{\theta}_T) - \nabla_i \mathcal{L}_T(\theta_0) = \{\nabla^2 \mathcal{L}_T(\theta_T^{(i)})(\hat{\theta}_T - \theta_0)\}_i$$

with $|\theta_T^{(i)} - \theta_0| \leq |\hat{\theta}_T - \theta_0| (i = 1, ..., p)$. If $\hat{\theta}_T$ lies in the interior of Θ we have $\nabla \mathcal{L}_T(\hat{\theta}_T) = 0$. If $\hat{\theta}_T$ lies on the boundary of Θ, then the assumption that θ_0 is in the interior implies $|\hat{\theta}_T - \theta_0| \geq \delta$ for some $\delta > 0$, i.e. we obtain $P(\sqrt{T}|\nabla \mathcal{L}_T(\hat{\theta}_T)| \geq \varepsilon) \leq P(|\hat{\theta}_T - \theta_0| \geq \delta) \to 0$ for all $\varepsilon > 0$. Thus, the result follows if we prove

$$\text{(i) } \nabla^2 \mathcal{L}_T(\theta_T^{(i)}) - \nabla^2 \mathcal{L}_T(\theta_0) \xrightarrow{P} 0$$

$$\text{(ii) } \nabla^2 \mathcal{L}_T(\theta_0) \xrightarrow{P} \Gamma$$

$$\text{(iii) } \sqrt{T} \nabla \mathcal{L}_T(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V).$$

We now obtain from (3.11)

$$\begin{aligned}
\nabla_{ij}^2 \mathcal{L}_T(\theta) \quad = \quad & -\frac{1}{4\pi} \frac{1}{T} \sum_{t=1}^{T} \int_{-\pi}^{\pi} \operatorname{tr} \left\{ f_\theta(\frac{t}{T}, \lambda) \nabla_{ij}^2 f_\theta(\frac{t}{T}, \lambda)^{-1} \right\} d\lambda \\
& -\frac{1}{4\pi} \frac{1}{T} \sum_{t=1}^{T} \int_{-\pi}^{\pi} \operatorname{tr} \left\{ \nabla_i f_\theta(\frac{t}{T}, \lambda) \nabla_j f_\theta(\frac{t}{T}, \lambda)^{-1} \right\} d\lambda \\
& + \frac{1}{8\pi^2 T} (\underline{X} - \underline{\mu}_\theta)' U_T(\nabla_{ij}^2 f_\theta^{-1}) (\underline{X} - \underline{\mu}_\theta) \\
& + \frac{1}{4\pi^2 T} (\nabla_i \underline{\mu}_\theta)' U_T(f_\theta^{-1}) (\nabla_j \underline{\mu}_\theta) \\
& - \frac{1}{4\pi^2 T} (\nabla_i \underline{\mu}_\theta)' U_T(\nabla_j f_\theta^{-1}) (\underline{X} - \underline{\mu}_\theta) \\
& - \frac{1}{4\pi^2 T} (\nabla_j \underline{\mu}_\theta)' U_T(\nabla_i f_\theta^{-1}) (\underline{X} - \underline{\mu}_\theta) \\
& - \frac{1}{4\pi^2 T} (\nabla_{ij}^2 \underline{\mu}_\theta)' U_T(f_\theta^{-1}) (\underline{X} - \underline{\mu}_\theta).
\end{aligned} \tag{3.13}$$

To prove (i) we have to consider the above terms separately. The assertion is obvious for the first and second term. Let $\theta_T = \theta_T^{(i)}$. The remaining terms of (3.13) can all be written as sums of expressions of the form

$$\frac{1}{T} \underline{X}' U_\theta \underline{X}, \quad \frac{1}{T} \underline{\nu}_\theta' U_\theta \underline{X} \quad \text{or} \quad \frac{1}{T} \underline{\nu}_{1\theta}' U_\theta \underline{\nu}_{2\theta} \tag{3.14}$$

with U_θ being equal to $U_T(f_\theta^{-1})$, $U_T(\nabla_i f_\theta^{-1})$ or $U_T(\nabla_{ij}^2 f_\theta^{-1})$. Lemma A.5(iii) implies $\|U_{\theta_T} - U_{\theta_0}\| \to 0$ in probability. Furthermore, $\frac{1}{T}\|\underline{\nu}_{\theta_T} - \underline{\nu}_{\theta_0}\|_2^2 \to 0$ in probability. This implies for example with the Cauchy-Schwarz inequality

$$
\begin{aligned}
&|\frac{1}{T}\underline{\nu}_{\theta_T}^{'}A_{\theta_T}\underline{X}-\frac{1}{T}\underline{\nu}_{\theta_0}^{'}A_{\theta_0}\underline{X}| \\
&\quad \leq \frac{1}{T}|(\underline{\nu}_{\theta_T}-\underline{\nu}_{\theta_0})'A_{\theta_T}\underline{X}|+\frac{1}{T}|\underline{\nu}_{\theta_0}^{'}(A_{\theta_T}-A_{\theta_0})\underline{X}| \\
&\quad \leq \frac{1}{T}\left\{\|\underline{\nu}_{\theta_T}-\underline{\nu}_{\theta 0}\|_2^2\|\underline{X}\|_2^2\right\}^{1/2}\|A_{\theta_T}\|+\frac{1}{T}\left\{\|\underline{\nu}_{\theta_0}\|_2^2\|\underline{X}\|_2^2\right\}^{1/2}\|A_{\theta_T}-A_{\theta_0}\|.
\end{aligned}
$$

As in the proof of Theorem 3.7 we have that $\frac{1}{T}\|\underline{X}\|_2^2$ is bounded in probability. Furthermore, $\|A_{\theta_T}\|$ is uniformly bounded by Lemma A.5 (iii). Therefore, the above expression tends to zero in probability. The other two expressions of (3.14) can be handled similarly which implies (i).

(ii) It follows from (3.12) (or from (3.13))

$$
\begin{aligned}
\nabla_{ij}^2\mathcal{L}_T(\theta_o) &= \frac{1}{8\pi^2 T}(\underline{X}-\underline{\mu})'U_T(\nabla_{ij}^2 f_\theta^{-1})(\underline{X}-\underline{\mu}) \\
&\quad +\frac{1}{4\pi^2 T}\nabla_{ij}^2\{(\underline{\mu}-\underline{\mu}_\theta)'U_T(f_\theta^{-1})\}(\underline{X}-\underline{\mu})+\text{ const.}
\end{aligned}
$$

and therefore (note that $\underline{X}$ is Gaussian)

$$
\begin{aligned}
\mathrm{var}(\nabla_{ij}^2\mathcal{L}_T(\theta_o)) &= \frac{1}{32\pi^4 T^2}\mathrm{tr}\{U_T(\nabla_{ij}^2 f_{\theta_o}^{-1})\Sigma U_T(\nabla_{ij}^2 f_{\theta_o}^{-1})\Sigma\} \\
&\quad +\frac{1}{16\pi^4 T^2}[\nabla_{ij}^2\{(\mu-\mu_{\theta_o})'U_T(f_{\theta_o}^{-1})\}]\Sigma[\nabla_{ij}^2\{U_T(f_{\theta_o}^{-1})(\mu-\mu_{\theta_o})\}].
\end{aligned}
$$

Lemma A.7 shows that this is of order $O(T^{-1})$. To calculate $E\nabla_{ij}^2\mathcal{L}_T(\theta_0)$ we consider again all terms of (3.13) separately. The expectation of the third term of (3.13) is

$$
\frac{1}{8\pi^2 T}\mathrm{tr}\left\{\Sigma U_T(\nabla_{ij}^2 f_\theta^{-1})\right\}+\frac{1}{8\pi^2 T}(\underline{\mu}-\underline{\mu}_\theta)'U_T(\nabla_{ij}^2 f_\theta^{-1})(\underline{\mu}-\underline{\mu}_\theta). \tag{3.15}
$$

The first two terms of (3.13) together with the first term of (3.15) tend with Lemma A.7 to the first and second term of Γ_{ij} and the expectation of the last four terms of (3.13) and the last term of (3.15) converge with Lemma A.7 to the last term of Γ_{ij} which proves (ii).

(iii) We use the method of cumulants. We have

$$
\begin{aligned}
0 = \nabla_i \mathcal{L}(\theta_0) \quad &= \quad \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \operatorname{tr} \left\{ (f - f_\theta) \nabla_i f_\theta^{-1} \right\} d\lambda du \\
&\quad + \frac{1}{4\pi} \int_0^1 \nabla_i (\mu_\theta(u) - \mu(u))' f_\theta^{-1}(u,0)(\mu_\theta(u) - \mu(u)) du.
\end{aligned}
$$

It follows from (3.11) and Lemma A.7 that $E\nabla_i \mathcal{L}_T(\theta_0)$ converges to the same expression with rate $O(T^{-1} \ln^5 T)$, i.e. we have

$$
\sqrt{T} E \nabla \mathcal{L}_T(\theta_0) = o(1).
$$

Furthermore, we get from (3.12)

$$
\begin{aligned}
& T \operatorname{cov}(\nabla_i \mathcal{L}(\theta_0),\ \nabla_j \mathcal{L}_T(\theta_0)) \\
& \quad = \quad \frac{1}{32\pi^4 T} \operatorname{tr} \left\{ U_T(\nabla_i f_{\theta_0}^{-1}) \Sigma U_T(\nabla_j f_{\theta_0}^{-1}) \Sigma \right\} \\
& \qquad\quad + \frac{1}{16\pi^4 T} \left[\nabla_i \left\{ (\underline{\mu} - \underline{\mu}_{\theta_0})' U_T(f_{\theta_0}^{-1}) \right\} \right] \Sigma \left[\nabla_j \left\{ U_T(f_{\theta_0}^{-1})(\underline{\mu} - \underline{\mu}_{\theta_0}) \right\} \right].
\end{aligned}
$$

Lemma A.7 implies that this tends to V_{ij}.

To study the higher-order cumulants we see from (3.12) that $\nabla_i \mathcal{L}_T(\theta_0)$ can be written as

$$
\nabla_i \mathcal{L}_T(\theta_0) = \frac{1}{8\pi^2 T} \underline{Y}' A_i \underline{Y} + \frac{1}{4\pi^2 T} \underline{\nu}_i' B \underline{Y} + \text{const.}
$$

where $E\underline{Y} = 0$. The cumulants of order ≥ 3 of the $\underline{\nu}_i B \underline{Y}$-terms are zero, while the mixed cumulants of the $\underline{Y}' A_i \underline{Y}$ and $\underline{\nu}_i' B \underline{Y}$-terms are nonzero if and only if there are exactly two $\underline{\nu}_i' B \underline{Y}$-terms involved (this follows from the product theorem for cumulants, cf. Brillinger, 1981, Theorem 2.3.2, $E\underline{Y} = 0$, and the normality of $\underline{Y}$).

Therefore, we obtain with the product theorem for cumulants

$$\begin{aligned} & T^{\ell/2}\mathrm{cum}(\nabla_{i_1}\mathcal{L}_T(\theta_0), ..., \nabla_{i_\ell}\mathcal{L}_T(\theta_0)) \\ = & \; C_1 T^{-\ell/2} \sum_{\substack{(j_1,...,j_\ell) \\ \text{permutation of} \\ (i_1,...,i_\ell)}} \mathrm{tr}\left\{\prod_{k=1}^{\ell} \Sigma\, U_T(\nabla_{j_k} f_\theta^{-1})\right\} \\ & + C_2 T^{-\ell/2} \sum_{\substack{(j_1,...,j_\ell) \\ \text{permutation of} \\ (i_1,...,i_\ell)}} \underline{\nu}'_{j_1} B \left\{\prod_{k=2}^{\ell-1} \Sigma\, U_T(\nabla_{j_k} f_\theta^{-1})\right\} \Sigma B \underline{\nu}'_{j_\ell}. \end{aligned}$$

Lemma A.1 implies that all terms are of order $O(T^{-\ell/2+1})$. Therefore, the theorem is proved. □

(3.9) Remark (i) Theorem 3.8 contains the asymptotic distribution of the Whittle-estimate in the stationary case as a special case (if f, f_{θ_o}, μ and μ_{θ_o} do not depend on u). The result for the classical Whittle-estimator is obtained if in addition $\mu = \mu_\theta = 0$ and $f = f_{\theta_o}$. Theorem 3.8 also gives the asymptotic distribution in the case where a stationary model is used with the classical Whittle-likelihood but the process is only locally stationary.
(ii) The matrices Γ and V from Theorem 3.8 simplify in several situations, in particular when the model is correctly specified ($f = f_{\theta_o}$, $\mu = \mu_{\theta_o}$ - cf. Remark 3.11 below), when a stationary model is fitted (f_θ and μ_θ do not depend on u), and when the parameters separate. For univariate processes this has been discussed in Dahlhaus (1996b, Remark 2.6 and 2.7) in the context of univariate maximum likelihood estimation.

The technical results proved in the appendix also can be used to derive the asymptotic properties of the exact maximum likelihood estimator

$$\tilde{\theta}_T := \operatorname*{argmin}_{\theta\in\Theta} \mathcal{L}_T^{(e)}(\theta).$$

(3.10) Theorem *Suppose that Assumption 3.6 holds. Then we have*

$$\sqrt{T}(\tilde{\theta}_T - \theta_o) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma^{-1}V\Gamma^{-1})$$

with Γ *and* V *as in Theorem 3.8.*

Proof. By using Lemma A.5 (ii) and Lemma A.8 (with $V_\ell = \Gamma_\ell^{-1}$ for all ℓ) the result can be proved in the same way as the result of Theorem 3.8 (cp. also Dahlhaus, 1996b, Theorem 2.4, where the asymptotic normality of the univariate maximum likelihood estimator has been proved by a slightly different proof method). Note, that first the consistency of the estimate has to be established. We omit details.

(3.11) Remark In the correctly specified case ($f = f_{\theta_o}$, $\mu = \mu_{\theta_o}$) it is easy to see that $V = \Gamma$ with

$$\Gamma_{ij} = \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \mathrm{tr}\left\{f_{\theta_o}(\nabla_i f_{\theta_o}^{-1}) f_{\theta_o}(\nabla_j f_{\theta_o}^{-1})\right\} d\lambda du + \frac{1}{2\pi}\int_0^1 (\nabla_i \mu_{\theta_o}(u))' f_{\theta_o}^{-1}(u,0)(\nabla_j \mu_{\theta_o}(u)) du.$$

In that case both estimates are asymptotically efficient. One way to see this is to prove an LAN-expansion and to show that $\sqrt{T}(\hat{\theta}_T - \theta_o)$ and $\sqrt{T}(\tilde{\theta}_T - \theta_o)$ are equivalent to the central sequence. For univariate processes and the MLE $\tilde{\theta}_T$ this has been done in Dahlhaus (1996b, Theorem 4.1 and 4.2). By using the technical lemmata of this appendix the LAN-property and the efficiency of both estimates can be derived in the same way as in that paper. We omit details.

A Appendix: Norms and matrix products of generalized Toeplitz matrices

In this section we study the behaviour of the matrix $U_T(\phi)$ in some detail. In particular, we prove that $U_T(\{4\pi^2 f\}^{-1})$ with $f(u,\lambda) = A(u,\lambda)A(u,-\lambda)'$ is a reasonable approximation of the inverse of $\Sigma_T(A,A)$. The results of this section are frequently used in Section 3. There are a few similarities to Section 4 of Dahlhaus (1996a) where we have constructed a different (less precise) approximation of the inverse of $\Sigma_T(A,A)$.

Suppose A is an $n \times n$ matrix. We denote by

$$\begin{aligned} \|A\| &= \sup_{x \in \mathbb{C}^n} \frac{|Ax|}{|x|} = \sup_{x \in \mathbb{C}^n} \left(\frac{x^* A^* A x}{x^* x} \right)^{1/2} \\ &= [\text{maximum characteristic root of } A^* A]^{1/2}, \end{aligned} \tag{A.1}$$

where A^* denotes the conjugate transpose of A, the spectral norm and by

$$\left\bracevert A \right\bracevert = [\mathrm{tr}(AA^*)]^{1/2} \tag{A.2}$$

the Euclidean norm of A. If A is a real positive semidefinite symmetric matrix, i.e. $A = P'DP$ with $PP' = P'P = I$ and $D = \mathrm{diag}\{\lambda_1, ..., \lambda_n\}$, where $\lambda_i \geq 0$, then we define $A^{1/2} = P'D^{1/2}P$, where $D^{1/2} = \mathrm{diag}\{\sqrt{\lambda_1}, ..., \sqrt{\lambda_n}\}$. Thus, $A^{1/2}$ is also positive semidefinite and symmetric with $A^{1/2}A^{1/2} = A$. Furthermore, $A^{-1/2} = (A^{1/2})^{-1}$ if A is positive definite.

The following results are well known [see, e.g., Davies (1973), Appendix II, or Graybill (1983), Section 5.6].

(A.1) Lemma *Let A, B be $n \times n$ matrices. Then*

(a) $|\mathrm{tr}(AB)| \leq \left\bracevert A \right\bracevert \left\bracevert B \right\bracevert$,

(b) $\left\bracevert AB \right\bracevert \leq \|A\| \left\bracevert B \right\bracevert$,

(c) $\left\bracevert AB \right\bracevert \leq \left\bracevert A \right\bracevert \|B\|$,

(d) $\|A\| \leq \left\bracevert A \right\bracevert < \sqrt{n}\|A\|$,

(e) $\|AB\| \leq \|A\|\|B\|$,

(f) $\|A\|^2 \leq (\sup_i \sum_{j=1}^n |a_{ij}|)(\sup_j \sum_{i=1}^n |a_{ij}|)$,

(g) $\|A\| = \sup_{x \in \mathbb{C}^n} \left| \frac{x^* A x}{x^* x} \right|$ *for A symmetric,*

(h) $|x^* A x| \leq x^* x \|A\|$, $x \in \mathbb{C}^n$,

(i) $\log \det A \leq \mathrm{tr}\{A - I\}$ *for A positive definite.*

Suppose now that the elements of A are continuously differentiable functions of θ. Then

(j) $\frac{\partial}{\partial \theta} A^{-1} = -A^{-1} \left(\frac{\partial}{\partial \theta} A \right) A^{-1}$,

(k) $\frac{\partial}{\partial \theta} \log \det A = \mathrm{tr} \left\{ A^{-1} \frac{\partial}{\partial \theta} A \right\}$.

Furthermore, let $L_T : \mathbb{R} \to \mathbb{R}$, $T \in \mathbb{R}^+$, be the periodic extension (with period 2π) of

$$L_T^*(\alpha) := \begin{cases} T, |\alpha| \le 1/T \\ 1/|\alpha|, 1/T \le |\alpha| \le \pi. \end{cases}$$

Properties of $L_T(\alpha)$ are listed in Dahlhaus (1997, Lemma A.4). We remark that we have with a generic constant

$$\int_{-\pi}^{\pi} L_T(\alpha) d\alpha \text{ is monotone increasing in } T, \tag{A.3}$$

$$L_T(\alpha) \le 2L_T(2\alpha), \tag{A.4}$$

$$\int_{-\pi}^{\pi} L_T(\beta - \alpha) L_T(\alpha + \gamma) d\alpha \le K L_T(\beta + \gamma) lnT, \tag{A.5}$$

$$\int_{-\pi}^{\pi} L_T(\alpha)^k d\alpha \le K T^{k-1} lnT^{\{k=1\}}. \tag{A.6}$$

Let

$$\Delta_T(\lambda) := \sum_{r=1}^{T} \exp(-i\lambda r).$$

Direct verification shows

$$|\Delta_T(\lambda)| \le \pi L_T(\lambda). \tag{A.7}$$

(A.2) Lemma *(i) Let $\psi : [0,1] \to \mathbb{C}$ be differentiable with bounded derivative. Then*

$$\begin{aligned} \sum_{r=1}^{T} \psi\left(\frac{r}{T}\right) \exp(-i\lambda r) &= \psi(1)\Delta_T(\lambda) + O(\sup_u |\psi'(u)| L_T(\lambda)) \\ &= O(L_T(\lambda)). \end{aligned}$$

The same holds if $\psi(\frac{r}{T})$ is replaced on the left side by $\psi_{r,T}$ with $\sup_r |\psi_{r,T} - \psi(\frac{r}{T})| = O(T^{-1})$.

(ii) Suppose $\psi : [0,1]^k \to \mathbb{C}$ *has bounded derivative* $\frac{\partial^k \psi}{\partial u_1 \ldots \partial u_k}$. *Then*

$$\left| \sum_{r_1,\ldots,r_k=1}^{T} \psi\left(\frac{r_1}{T},\ldots,\frac{r_k}{T}\right) \exp\left(-i\sum_{j=1}^{k}\lambda_j r_j\right)\right|$$
$$\leq K \sup_{\ell \leq k} \sup_{\{i_1,\ldots,i_\ell\}\subset\{1,\ldots k\}} \sup_u \left|\frac{\partial^\ell}{\partial u_{i_1}\ldots\partial u_{i_\ell}}\psi(u)\right| \prod_{j=1}^{k} L_T(\lambda_j) = O\left(\prod_{j=1}^{k} L_T(\lambda_j)\right).$$

Proof. (i) Summation by parts gives

$$\sum_{r=1}^{T}\psi(\frac{r}{T})\exp(-i\lambda r) = -\sum_{r=1}^{T-1}\left\{\psi(\frac{r+1}{T}) - \psi(\frac{r}{T})\right\}\Delta_r(\lambda) + \psi(1)\Delta_T(\lambda)$$

which implies with (A.7) the result. (ii) Let D_j be the difference operator with respect to the $j-th$ component, i.e. $D_j\psi(\frac{r_1}{T},\ldots,\frac{r_k}{T}) := \psi(\frac{r_1}{T},\ldots,\frac{r_{j-1}}{T},\frac{r_j+1}{T},\frac{r_{j+1}}{T},\ldots,\frac{r_k}{T}) - \psi(\frac{r_1}{T},\ldots,\frac{r_k}{T})$. Then we obtain with repeated partial summation and the convention $\psi(u) = 0$ for $u \notin [0,1]^k$

$$\sum_{r_1,\ldots,r_k=1}^{T}\psi\left(\frac{r_1}{T},\ldots,\frac{r_k}{T}\right)\exp\left(-i\sum_{j=1}^{k}\lambda_j r_j\right)$$
$$= (-1)^k \sum_{r_1,\ldots,r_k=1}^{T}\left(D_1\ldots D_k\psi\left(\frac{r_1}{T},\ldots,\frac{r_k}{T}\right)\right)\prod_{j=1}^{k}\Delta_{r_j}(\lambda_j).$$

We have

$$\left|D_1\ldots D_k\psi\left(\frac{r_1}{T},\ldots,\frac{r_k}{T}\right)\right| \leq 2^{k-\ell}T^{-\ell}\sup_u\left|\frac{\partial^\ell}{\partial u_{i_1}\ldots\partial u_{i_\ell}}\psi(u)\right|$$

where $\{i_1,\ldots,i_\ell\} = \{i | r_i \neq T\}$, leading to the result. □

(A.3) Assumption

(i) Suppose $A : [0,1]\times\mathbb{R} \to \mathbb{C}^{d\times d}$ is a 2π-periodic matrix function with $A(u,\lambda) = \overline{A(u,-\lambda)}$ whose components are differentiable in u and λ with uniformly bounded derivatives $\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda}A_{ab}$. $A^o_{t,T} : \mathbb{R} \to \mathbb{C}^{d\times d}$ are 2π-periodic matrix functions with $\sup_{t,\lambda}|A^o_{t,T}(\lambda)_{ab} - A(\frac{t}{T},\lambda)_{ab}| \leq KT^{-1}$ for all $a,b \in \{1,\ldots,d\}$.

(ii) Suppose in addition to (i) that all eigenvalues of $A(u,\lambda)\overline{A(u,\lambda)}'$ are bounded from below by some $C > 0$ uniformly in u and λ.

(iii) Suppose $\phi : [0,1] \times \mathbb{R} \to \mathbb{C}^{d\times d}$ is a 2π-periodic matrix function whose components are twice differentiable in u and differentiable in λ with uniformly bounded derivative $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial \lambda}\phi$.

(iv) Suppose the components of $\mu : [0,1] \to \mathbb{R}^d$ are differentiable with uniformly bounded derivatives.

(A.4) Remark All results stated in this appendix are uniform in the sense that the upper bounds depend only on the bounds of the involved functions A, ϕ and μ and their derivatives and not on the particular values.

(A.5) Lemma

(i) Suppose A and B fulfill Assumption A.3 (i) and the components of ϕ are differentiable with uniformly bounded derivative $\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda}\phi_{ab}$. Then we have

$$\|\Sigma_T(A,B)\| \le C_1$$

and

$$\|U_T(\phi)\| \le C_2$$

with some constants C_1, C_2.

(ii) More precisely we have under Assumption A.3 (i)

$$\|\Sigma_T(A,A)\| \le 2\pi \sup_{u,\lambda} \|A(u,\lambda)\overline{A(u,\lambda)}'\| + C_A o(1).$$

where C_A is a constant depending on the upper bounds of A and its derivatives. If in addition A fulfills Assumption A.3 (ii) we have

$$\|\Sigma_T(A,A)^{-1}\| \le (2\pi \inf_{u,\lambda} \lambda_{\min}^{|A|^2}(u,\lambda) + C_A o(1))^{-1}$$

where $\lambda_{\min}^{|A|^2}(u,\lambda)$ is the smallest eigenvalue of $A(u,\lambda)\overline{A(u,\lambda)}'$.

(iii) If ϕ is symmetric and fulfills Assumption A.3 (iii) we have

$$\|U_T(\phi)\| \leq 2\pi \sup_{u,\lambda} \|\phi(u,\lambda)\| + C_\phi o(1)$$

where C_ϕ is a constant depending on the upper bounds of ϕ and its derivatives. If in addition the smallest eigenvalue $\lambda^\phi_{\min}(u,\lambda)$ of $\phi(u,\lambda)$ is uniformly bounded from below, then

$$\|U_T(\phi)^{-1}\| \leq (2\pi \inf_{u,\lambda} \lambda^\phi_{\min}(u,\lambda) + C_\phi o(1))^{-1}.$$

Proof. (i) Lemma A.1 (f) implies

$$\|U_T(\phi)\| \leq d \sum_{r\in\mathbb{Z}} \sup_{\substack{u\in[0,1] \\ a,b \in \{1,\ldots,d\}}} \left| \int \phi(u,\lambda)_{ab} \exp(i\lambda r) d\lambda \right| + K.$$

The smoothness conditions then imply the result (cf. Dahlhaus, 1996a, p. 156). The upper bound for $\|\Sigma_T(A,B)\|$ is obtained in the same way.
(ii) follows for $d=1$ from Lemma 4.4 of Dahlhaus (1996a). In the multivariate case the proof is completely analogous to that lemma. We omit the details.
(iii) The bounds for $\|U_T(\phi)\|$ and $\|U_T(\phi)^{-1}\|$ can be established in exactly the same way as the bounds for $\|\Sigma_T(A,A)\|$ and $\|\Sigma_T(A,A)^{-1}\|$ by a straightforward generalisation of Lemma 4.4 of Dahlhaus (1996a). We omit the details. □

In the proof of Lemma A.7 we frequently make use of the following result.

(A.6) Lemma *Suppose A and B fulfill Assumption A.3 (i) and ϕ fulfills Assumption A.3 (iii) with $d=1$. Then we have*

$$\begin{aligned}
&\sum_{r,s=1}^{T} \phi\left(\frac{1}{T}\left[\frac{r+s}{2}\right]^*, \lambda\right) A^o_{s,T}(\gamma_1) B^o_{r,T}(-\gamma_2) \exp\{-i(\lambda-\gamma_1)s - i(\gamma_2-\lambda)r\} \\
&= \sum_{r,s=1}^{T} \phi\left(\frac{r+s}{2T}, \lambda\right) A\left(\frac{s}{T}, \lambda\right) B\left(\frac{r}{T}, -\lambda\right) \exp\{-i(\lambda-\gamma_1)s - i(\gamma_2-\lambda)r\} \\
&\qquad + O(L_T(2\lambda - 2\gamma_1)) + O(L_T(2\gamma_2 - 2\lambda)) \\
&= O(L_T(2\lambda-2\gamma_1)L_T(2\gamma_2-2\lambda))
\end{aligned}$$

PROOF. We start by replacing $A^o_{s,T}(\gamma_1)$ by $A(\frac{s}{T},\gamma_1)$. Lemma A.2 (i) and (A.4) imply

$$\left|\sum_{r=1}^{T}\phi\left(\frac{1}{T}\left[\frac{r+s}{2}\right]^*,\lambda\right)B^o_{r,T}(-\gamma_2)\exp\{-i(\gamma_2-\lambda)r\}\right|\le KL_T(2\gamma_2-2\lambda)$$

which gives a replacement error of $KL_T(2\gamma_2-2\lambda)$. In the same way we replace $B^o_{r,T}(-\gamma_2)$ by $B(\frac{r}{T},-\gamma_2)$. We then replace $\phi(\frac{1}{T}[\frac{r+s}{2}]^*,\lambda)$ by $\phi(\frac{r+s}{2T},\lambda)$. For $r+s$ even those two are the same. The replacement error therefore is ($r=2k, s=2\ell-1$)

$$\sum_{k,\ell=1}^{[T/2]}\left[\phi\left(\frac{2(k+\ell)}{2T},\lambda\right)-\phi\left(\frac{2(k+\ell)-1}{2T},\lambda\right)\right]A\left(\frac{2\ell-1}{T},\gamma_1\right)B\left(\frac{2k}{T},-\gamma_2\right)\cdot$$
$$\cdot\exp\{-i(\lambda-\gamma_1)(2\ell-1)-i(\gamma_2-\lambda)2k\}+\text{a similar term.}$$

Since $\phi(\frac{2(k+\ell)}{2T},\lambda)-\phi(\frac{2(k+\ell)-1}{2T},\lambda)=\frac{1}{T}[\frac{\partial}{\partial u}\phi(\frac{k+\ell}{T},\lambda)+O(T^{-1})]$ we get with Lemma A.2 (i) that this expression is bounded by $KL_T(2\lambda-2\gamma_1)$. Finally, we replace $A(\frac{s}{T},\gamma_1)$ by $A(\frac{s}{T},\lambda)$ with a replacement error of $K|\lambda-\gamma_1\ |L_T(\lambda-\gamma_1)L_T(\gamma_2-\lambda)\le KL_T(2\gamma_2-2\lambda)$ (by using Lemma A.2 (ii)). Similarly, we obtain $KL_T(2\lambda-2\gamma_1)$ as the replacement error for replacing $B(\frac{r}{T},-\gamma_2)$ by $B(\frac{r}{T},-\lambda)$ which leads to the first equation. The second equation then follows with Lemma A.2 (ii) and (A.4). □

(A.7) Lemma *Let $k\in\mathbb{N}$, A_ℓ, B_ℓ fulfill Assumption A.3 (i), ϕ_ℓ fulfill Assumption A.3 (iii) and μ_1,μ_2 fulfill Assumption A.3 (iv). Then we have*

(i) $\frac{1}{T}\mathrm{tr}\{\prod_{\ell=1}^{k}U_T(\phi_\ell)\Sigma_T(A_\ell,B_\ell)\}$

$$=(2\pi)^{2k-1}\int_0^1\int_{-\pi}^{\pi}\mathrm{tr}\{\prod_{\ell=1}^{k}\phi_\ell(u,\lambda)A_\ell(u,\lambda)B_\ell(u,-\lambda)'\}d\lambda du+O(T^{-1}\ln^{2k-1}T),$$

(ii) $\frac{1}{T}\mu'_{1T}\{\prod_{\ell=1}^{k-1}U_T(\phi_\ell)\Sigma_T(A_\ell,B_\ell)\}U_T(\phi_k)\mu_{2T}$

$$=(2\pi)^{2k-1}\int_0^1\mu_1(u)'\{\prod_{\ell=1}^{k-1}\phi_\ell(u,0)A_\ell(u,0)B_\ell(u,0)'\}\phi_k(u,0)\mu_2(u)du+O(T^{-1}\ln^{2k-1}T).$$

Remark. If $\tilde{I}$ is the $d \times d$ identity matrix then $\frac{1}{2\pi}\Sigma_T(\tilde{I},\tilde{I}) = \frac{1}{2\pi}U_T(\tilde{I})$ is the $dT \times dT$ identity matrix. Therefore Lemma A.7 also give the asymptotic expressions for

$$\frac{1}{T}\mathrm{tr}\left\{\prod_{\ell=1}^{k}\Sigma_T(A_\ell,B_\ell)\right\} \qquad \text{and} \qquad \frac{1}{T}\mathrm{tr}\left\{\prod_{\ell=1}^{k}U_T(\phi_\ell)\right\}$$

and more general for the trace of an arbitrary product of Σ_T's and U_T's.

PROOF. (i) We give the proof for $k=1$ and afterwards for general $k \geq 2$. We have

$$\begin{aligned}
&\frac{1}{T}\mathrm{tr}\{U_T(\phi)\Sigma_T(A,B)\} \\
&= \frac{1}{T}\sum_{a,b,c=1}^{d}\sum_{r,s=1}^{T}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\phi\left(\frac{1}{T}\left[\frac{r+s}{2}\right]^*,\lambda\right)_{ab}A^o_{s,T}(\gamma)_{bc}B^o_{r,T}(-\gamma)_{ac}\exp\{i(\lambda-\gamma)(r-s)\}d\lambda d\gamma
\end{aligned}$$

which by using Lemma A.6 and (A.6) is equal to

$$\begin{aligned}
&\frac{1}{T}\sum_{a,b,c=1}^{d}\sum_{r,s=1}^{T}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\phi\left(\frac{r+s}{2T},\lambda\right)_{ab}A\left(\frac{s}{T},\lambda\right)_{bc}B\left(\frac{r}{T},-\lambda\right)_{ac}\exp\{i(\lambda-\gamma)(r-s)\}d\lambda d\gamma \\
&\qquad +O(T^{-1}\ln T).
\end{aligned}$$

Integration over γ now gives the result.

To simplify notation we use in the rest of the proof the "trace"-notation keeping in mind that in the calculation of remainders usually the individual components have to be considered. For $k \geq 2$ we then have

$$\begin{aligned}
&\frac{1}{T}\mathrm{tr}\left\{\prod_{j=1}^{k}U_T(\phi_j)\Sigma_T(A_j,B_j)\right\} \\
&= \frac{1}{T}\sum_{\substack{r_1,\ldots,r_k,\\ s_1,\ldots,s_k=1}}^{T}\int_{-\pi}^{\pi}\!\!\cdots\!\int\mathrm{tr}\left\{\prod_{j=1}^{k}\phi_j\left(\frac{1}{T}\left[\frac{r_j+s_j}{2}\right]^*,\lambda_j\right)A^o_{j,s_j,T}(\gamma_j)B^o_{j,r_{j+1},T}(-\gamma_j)\right\} \\
&\qquad \times\exp\left\{-i\sum_{j=1}^{k}[(\lambda_j-\gamma_j)s_j+(\gamma_j-\lambda_{j+1})r_{j+1}]\right\}d\lambda d\gamma
\end{aligned}$$

where $r_{k+1}=r_1$ and $\lambda_{k+1}=\lambda_1$. Application of Lemma A.6 together with (A.5) and (A.6) shows that this is equal to

$$\frac{1}{T}\sum_{\substack{r_1,\ldots,\;r_k,\\ s_1,\ldots,\;s_k=1}}^{T}\int_{-\pi}^{\pi}\!\!...\!\int \operatorname{tr}\left\{\prod_{j=1}^{k}\phi_j\left(\frac{r_j+s_j}{2T},\lambda_j\right)A_j\left(\frac{s_j}{T},\lambda_j\right)B_j\left(\frac{r_{j+1}}{T},-\lambda_{j+1}\right)\right\}$$

$$\times\exp\left\{-i\sum_{j=1}^{k}[(\lambda_j-\gamma_j)s_j+(\gamma_j-\lambda_{j+1})r_{j+1}]\right\}d\lambda d\gamma+O(T^{-1}\ln^{2k-1}T).$$

Integration over all γ_j shows that this is equal to

$$(2\pi)^k\frac{1}{T}\sum_{r_1,\ldots,r_k=1}^{T}\int_{-\pi}^{\pi}\!\!...\!\int \operatorname{tr}\left\{\prod_{j=1}^{k}\phi_j\left(\frac{r_j+r_{j+1}}{2T},\lambda_j\right)A_j\left(\frac{r_{j+1}}{T},\lambda_j\right)B_j\left(\frac{r_{j+1}}{T},-\lambda_{j+1}\right)\right\}$$

$$\times\exp\left\{-i\sum_{j=1}^{k}(\lambda_j-\lambda_{j+1})r_{j+1}\right\}d\lambda+O(T^{-1}\ln^{2k-1}T).$$

We now replace the argument λ_k in ϕ_k, A_k and B_{k-1} by λ_{k-1}. The replacement error is of the form

$$\frac{1}{T}\sum_{r_1,\ldots,r_k=1}^{T}\int_{-\pi}^{\pi}\!\!...\!\int \psi_\lambda\left(\frac{r_1}{T},\ldots,\frac{r_k}{T}\right)\exp\left\{-i\sum_{j=1}^{k}(\lambda_j-\lambda_{j+1})r_{j+1}\right\}d\lambda$$

where $\sup_u|\frac{\partial^\ell}{\partial u_{1_1}\ldots\partial u_{i_\ell}}\psi(u)|\le K|\lambda_k-\lambda_{k-1}|$ for all $\{i_1,\ldots,i_\ell\}\subset\{1,\ldots,k\}$, i.e. we obtain for the replacement error with Lemma A.2 (ii) and (A.6) as an upper bound

$$K\frac{1}{T}\int_{-\pi}^{\pi}\!\!...\!\int|\lambda_k-\lambda_{k-1}|\prod_{j=1}^{k}L_T(\lambda_j-\lambda_{j-1})d\lambda\le KT^{-1}\ln^{k-1}T.$$

In the same way we successively replace all λ_j by λ_1 and integrate finally over $\lambda_2,\ldots,\lambda_k$ which proves the assertion.

(ii) The proof of (ii) is completely analogous to (i). We therefore only give a brief sketch for the case $k \geq 2$. We have

$$
\begin{aligned}
&\frac{1}{T}\mu_{1T}^{'}\left\{\prod_{j=1}^{k-1} U_T(\phi_j)\Sigma_T(A_j,B_j)\right\} U_T(\phi_k)\mu_{2T} \\
&= \frac{1}{T}\sum_{\substack{r_1,\ldots,r_k,\\ s_1,\ldots,s_k=1}}^{T} \int_{-\pi}^{\pi}\!\!\ldots\!\int \mu_1(\frac{r_1}{T})'\left\{\prod_{j=1}^{k-1}\phi_j\left(\frac{1}{T}\left[\frac{r_j+s_j}{2}\right]^*,\lambda_j\right)A^o_{j,s_j,T}(\gamma_j)B^o_{j,r_{j+1},T}(-\gamma_j)\right\} \\
&\quad\times\phi_k\left(\frac{1}{T}\left[\frac{r_k+s_k}{2}\right]^*,\lambda_k\right)\mu_2\left(\frac{s_k}{T}\right) \\
&\quad\times\exp\left\{-i\sum_{j=1}^{k-1}[(\lambda_j-\gamma_j)s_j+(\gamma_j-\lambda_{j+1})r_{j+1}]+i\lambda_1 r_1-i\lambda_k s_k\right\}d\lambda d\gamma.
\end{aligned}
$$

We now use similar replacement steps as in (i) (note that Lemma A.6 also holds if e.g. $B^o_{r,T}(-\gamma_2)=\mu_1(\frac{r}{T})$ and γ_2 is set equal to zero) which leads with $s_k=r_{k+1}$ to

$$
\begin{aligned}
&(2\pi)^{k-1}\frac{1}{T}\sum_{r_1,\ldots,r_{k+1}=1}^{T}\int_{-\pi}^{\pi}\!\!\ldots\!\int\mu_1\left(\frac{r_1}{T}\right)'\left\{\prod_{j=1}^{k-1}\phi_j\left(\frac{r_j+r_{j+1}}{2T},\lambda_j\right)A_j\left(\frac{r_{j+1}}{T},\lambda_j\right)B_j\left(\frac{r_{j+1}}{T},-\lambda_{j+1}\right)\right\} \\
&\quad\times\phi_k\left(\frac{r_k+r_{k+1}}{T},\lambda_k\right)\mu_2\left(\frac{r_{k+1}}{T}\right)\exp\left\{i\lambda_1 r_1-i\sum_{j=1}^{k-1}(\lambda_j-\lambda_{j+1})r_{j+1}-i\lambda_k r_{k+1}\right\}d\lambda \\
&\quad+O(T^{-1}\ln^{2k-1}T).
\end{aligned}
$$

As before we now replace all λ_j by λ_1 and finally λ_1 by 0 leading to the result. $\square$

(A.8) Lemma *Let $k\in\mathbb{N}$ and $\{I_1,\ldots,I_4\}$ be a partition of $\{1,\ldots,k\}$. Let the matrices A_ℓ, B_ℓ (for $\ell\in I_1$) fulfill Assumption A.3 (i), C_ℓ (for $\ell\in I_2$) fulfill Assumption A.3 (i), (ii) with bounded derivatives $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial\lambda}C_\ell(u,\lambda)_{ab}$, ϕ_ℓ (for $\ell\in I_3\cup I_4$) fulfill Assumption A.3 (iii) with eigenvalues (for $\ell\in I_4$) that are bounded from below uniformly in u and λ, and μ_1,μ_2 fulfill Assumption A.3 (iv).*

Let further

$$\begin{array}{lll} V_\ell = \Sigma_T(A_\ell, B_\ell), & \psi_\ell(u,\lambda) = 2\pi A_\ell(u,\lambda) B_\ell(u,-\lambda)' & (\ell \in I_1), \\ V_\ell = \Sigma_T(C_\ell, C_\ell)^{-1}, & \psi_\ell(u,\lambda) = \frac{1}{2\pi} C_\ell(u,-\lambda)'^{-1} C_\ell(u,\lambda)^{-1} & (\ell \in I_2), \\ V_\ell = U_T(\phi_\ell), & \psi_\ell(u,\lambda) = 2\pi \phi_\ell(u,\lambda) & (\ell \in I_3), \\ V_\ell = U_T(\phi_\ell)^{-1} & \psi_\ell(u,\lambda) = \frac{1}{2\pi} \phi_\ell(u,\lambda)^{-1} & (\ell \in I_4). \end{array}$$

Then we have

(i)

$$\frac{1}{T}\mathrm{tr}\left\{\prod_{\ell=1}^k V_\ell\right\} = \frac{1}{2\pi}\int_0^1\int_{-\pi}^{\pi} \mathrm{tr}\left\{\prod_{\ell=1}^k \psi_\ell(u,\lambda)\right\} d\lambda du + O(T^{-1} ln^{6k-1} T),$$

(ii)

$$\frac{1}{T}\mu_{1T}'\left\{\prod_{\ell=1}^k V_\ell\right\}\mu_{2T} = \frac{1}{2\pi}\int_0^1 \mu_1(u)'\left\{\prod_{\ell=1}^k \psi_\ell(u,0)\right\}\mu_2(u) du + O(T^{-1} ln^{6k-1} T).$$

PROOF. (i) Let $j = |I_2| + |I_4|$. More precisely we prove the result with the rate $O(T^{-1} ln^{2k+4j-1})$. For $j = 0$ the assertion follows for all k from Lemma A.7. Suppose now the assertion holds for all k and some fixed j. Consider the case $j+1$. By renumbering the V_ℓ we can assume that $k \in I_2 \cup I_4$. Suppose $k \in I_2$. We approximate $V_k = \Sigma^{-1} := \Sigma_T(C_k, C_k)^{-1}$ by $\tilde{U} := U_T(\{4\pi^2 C_k \bar{C}_k'\}^{-1})$. We have with Lemma A.1, Lemma A.5 and Proposition 3.3

$$\begin{aligned} & \left|\frac{1}{T}\mathrm{tr}\left\{\prod_{\ell=1}^k V_\ell\right\} - \frac{2}{T}\mathrm{tr}\left\{\left(\prod_{\ell=1}^{k-1} V_\ell\right)\tilde{U}\right\} + \frac{1}{T}\mathrm{tr}\left\{\left(\prod_{\ell=1}^{k-1} V_\ell\right)\tilde{U}\Sigma\tilde{U}\right\}\right| \\ = & \left|\frac{1}{T}\mathrm{tr}\left\{\left(\prod_{\ell=1}^{k-1} V_\ell\right)(\Sigma^{-1} - \tilde{U})\Sigma(\Sigma^{-1} - \tilde{U})\right\}\right| \\ \leq & \left(\prod_{\ell=1}^{k-1}\|V_\ell\|\right)\|\Sigma\|\frac{1}{T}\,\big|\!\big|\!\big|\,\Sigma^{-1} - \tilde{U}\,\big|\!\big|\!\big|^2 = O(T^{-1} ln^3 T). \end{aligned}$$

This implies the convergence with rate $O(T^{-1} ln^{2(k+2)+4j-1}) = O(T^{-1} ln^{2k+4(j+1)-1})$ which gives the result. If $k \in I_4$ the result is obtained in the same way by using the second equation of Proposition 3.3. (ii) follows similarly. □

Technically, Lemma A.7 and Lemma A.8 are the key results for proving the asymptotic properties of the local likelihood estimator and of the exact MLE as done in Section 3. For $I_1 = \{\ell|\ell \text{ even}\}$, $I_2 = \{\ell|\ell \text{ odd}\}$, $I_3 = I_4 = \emptyset$ Lemma A.8 is a generalisation of a central result for Gaussian stationary processes to the locally stationary case (cf. Taniguchi, 1983, Theorem 1). Note, that it is not very difficult to improve the rate $ln^{6k-1}T$ in the above lemma.

(A.9) Proof of Proposition 3.4 We replace $\Sigma_T := \Sigma_T(A, A)$ by $U_T := U_T(A\bar{A}')$. We obtain with Lemma A.1 (i)

$$\begin{aligned} |\frac{1}{T}\log\det\Sigma_T - \frac{1}{T}\log\det U_T| &= |\frac{1}{T}\log\det\Sigma_T^{-1/2}U_T\Sigma_T^{-1/2}| \\ &\leq \max\{\frac{1}{T}\mathrm{tr}(\Sigma_T^{-1}U_T - I), \frac{1}{T}\mathrm{tr}(U_T^{-1}\Sigma_T - I)\}. \end{aligned}$$

Lemma A.8 yields that both terms are of order $O(T^{-1}ln^{11}T)$. Since $f(u,\lambda) = A(u,\lambda)\overline{A(u,\lambda)}'$ is symmetric and positive definite there exist an orthonormal matrix $B(u,\lambda)$ and a diagonal matrix $D(u,\lambda) = \mathrm{diag}\{d_1(u,\lambda), \ldots, d_d(u,\lambda)\}$ with positive $d_j(u,\lambda)$ such that

$$f(u,\lambda) = B(u,\lambda)D(u,\lambda)B(u,\lambda)'.$$

Now let $x \in [0,1]$ and

$$f^{(x)}(u,\lambda) := B(u,\lambda)D^{(x)}(u,\lambda)B(u,\lambda)'$$

with

$$D^{(x)}(u,\lambda) := \mathrm{diag}\{d_1(u,\lambda)^x, \ldots, d_d(u,\lambda)^x\}.$$

We have $U_T(f^{(1)}) = U_T$ and $U_T(f^{(0)}) = 2\pi I$ where I in the $dT \times dT$ identity matrix. We therefore obtain with $U_T^{(x)} := U_T(f^{(x)})$

$$\begin{aligned} \frac{1}{T}\log\det\Sigma_T &= \frac{1}{T}\log\det U_T + O(T^{-1}ln^{11}T) \\ &= \frac{1}{T}\int_0^1 \frac{\partial}{\partial x}\log\det U_T^{(x)}dx + \log(2\pi)^d + O(T^{-1}ln^{11}T) \\ &= \frac{1}{T}\int_0^1 \mathrm{tr}\left[U_T^{(x)-1}\frac{\partial}{\partial x}U_T^{(x)}\right]dx + \log(2\pi)^d + O(T^{-1}ln^{11}T). \end{aligned}$$

Furthermore

$$\frac{\partial}{\partial x} U_T^{(x)} = \int_{-\pi}^{\pi} \exp(i\lambda(r-s)) \frac{\partial}{\partial x} f^{(x)}(\frac{1}{T}[\frac{r+s}{2}]^*, \lambda) d\lambda$$

with

$$\frac{\partial}{\partial x} f^{(x)}(u,\lambda) = B(u,\lambda) \text{diag}\{d_1(u,\lambda)^x \log d_1(u,\lambda), \ldots, d_d(u,\lambda)^x \log d_d(u,\lambda)\} B(u,\lambda)'.$$

Since $f^{(x)}(u,\lambda)$ and $\frac{\partial}{\partial x} f^{(x)}(u,\lambda)$ have the same smoothness properties as $\phi(u,\lambda)$ uniformly in x we obtain from Lemma A.8 with straightforward calculations

$$\begin{aligned} \frac{1}{T} \text{tr} \left[U_T^{(x)-1} \frac{\partial}{\partial x} U_T^{(x)} \right] &= \frac{1}{2\pi} \int_0^1 \int_{-\pi}^{\pi} \left\{ \sum_{j=1}^{d} \log d_j(u,\lambda) \right\} d\lambda du + O(T^{-1} \ln^{11} T) \\ &= \frac{1}{2\pi} \int_0^1 \int_{-\pi}^{\pi} \log \det f(u,\lambda) d\lambda du + O(T^{-1} \ln^{11} T) \end{aligned}$$

uniformly in x which implies the result. □

Acknowledgement. The author is grateful to M. Sahm for correcting a mistake in the proof of Proposition 3.4.

References

Brillinger, D.R. (1981) *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

Dahlhaus, R. (1996a) On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Proc. Appl.* **62**, 139-168.

Dahlhaus, R. (1996b) Maximum likelihood estimation and model selection for locally stationary processes. *J. Nonpar. Statist.* **6**, 171-191.

Dahlhaus, R. (1996c) Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In: *Athens Conference on Applied Probability and Time Series Vol II.* (P.M. Robinson and M. Rosenblatt, eds.), 145-159, Lecture Notes in Statistics 115, Springer, New York.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1-37.

Davies, R.B. (1973) Asymptotic inference in stationary Gaussian time series. *Adv. Appl. Prob.* **5**, 469-497.

Dunsmuir, W. (1979) A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Ann. Statist.* **7**, 490-506.

Dzhaparidze, K. (1971) On methods for obtaining asymptotically efficient spectral parameter estimates for a stationary Gaussian process with rational spectral density. *Theory Probab. Appl.* **16**, 550-554.

Dzhaparidze, K. (1986) *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.* Springer, New York.

Fox, R. and Taqqu, M.S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14**, 517-532.

Graybill, F.A. (1983) *Matrices with Applications in Statistics.* Wadsworth, Belmont, California.

Grenander, U. and Szegö, G. (1958) *Toeplitz Forms and their Applications.* University of California Press, Berkeley.

Hannan, E.J. (1973) The asymptotic theory of linear time series models. *J. Appl. Prob.* **10**, 130-145.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Hosoya, Y. and Taniguchi, M. (1982) A central limit theorem for stationary processes and the parameter estimation of linear processes *Ann. Statist.* **10**, 132-153.

Klüppelberg, C. and Mikosch, T. (1996) Parameter estimation for a misspecified ARMA process with infinite variance innovations. *J. Math. Sci* **78**, 60-65.

Mallat, S., Papanicolaou, G. and Zhang, Z. (1998) Adaptive covariance estimation of locally stationary processes. *Ann. Statist.* **26**, 1-47.

Neumann, M.H. and von Sachs, R. (1997) Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.* **25**, 38-76.

Parzen, E. (1983) Autoregressive spectral estimation. In: *Handbook of Statistics* (D.R. Brillinger and P.R. Krishnaiah, eds.), **3**, 221-247, North-Holland, Amsterdam.

Priestley, M.B. (1965) Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. Ser. B* **27**, 204-237.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*, Vol.**2**, Academic Press, London.

Robinson, P.M. (1995) Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23**, 1630-1661.

Taniguchi, M. (1983) On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann. Statist.* **11**, 157-169.

Walker, A.M. (1964) Asymptotic properties of least squares estimates of the parameters of the spectrum of a stationary non-deterministic time series. *J. Austral. Math. Soc.* **4**, 363-384.

Whittle, P. (1953) Estimation and information in stationary time series. *Ark. Mat.*, **2**, 423-434.

Whittle, P. (1954) Some recent contributions to the theory of stationary processes. Appendix to *A study in the analysis of stationary time series*, by H. Wold, 2nd ed. 196-228. Almqvist and Wiksell, Uppsala.